Research
Report

An Entropy-Based Measure of Dependence Between Two Groups of Random Variables

Nan Kong

Research &
Development

May 2007
RR-07-20

An Entropy-Based Measure of Dependence Between Two Groups of Random Variables

Nan Kong
ETS, Princeton, NJ

May 2007

Abstract

In multivariate statistics, the linear relationship among random variables has been fully explored in the past. This paper looks into the dependence of one group of random variables on another group of random variables using (conditional) entropy. A new measure, called the K-dependence coefficient or dependence coefficient, is defined using (conditional) entropy. This paper shows that the K-dependence coefficient is a measure of the degree of dependence of one group of random variables on another group of random variables. The dependence measured by the K-dependence coefficient includes both linear and nonlinear dependencies between two groups of random variables. Therefore, the K-dependence coefficient measures the total degree of dependence and not just the linear component of the dependence between the two groups of random variables. Furthermore, the concept of the K-dependence coefficient is extended by defining the partial K-dependence coefficient and the semipartial K-dependence coefficient. Properties of partial and semipartial K-dependence coefficients are also explored.

Key words: Entropy, conditional entropy, K-dependence coefficient, partial K-dependence coefficient, semipartial K-dependence coefficient

Acknowledgments

The author would like to express his sincere thanks to Professor Charles Lewis for his valuable suggestions and comments. His encouragement and support enabled the author to finish this paper. The author is also very grateful to Dan Eignor, Shelby Haberman, and Kim Fryer for their helpful comments and suggestions.

1. Introduction

Entropy is a measure of uncertainty for a random system or a group of random variables. Let $\vec{X} = (x_1, x_2, \ldots, x_n)^T$ and $\vec{Y} = (y_1, y_2, \ldots, y_m)^T$ be two groups of discrete random variables with N and M possible states, respectively, which are denoted by $\{\vec{x}_1, \ldots, \vec{x}_N\}$ and $\{\vec{y}_1, \ldots, \vec{y}_M\}$. Joint probabilities of these possible states are denoted by $p(\vec{x}_i, \vec{y}_j)$ $(i = 1, \ldots, N; j = 1, \ldots, M)$, while marginal probabilities are denoted by $p(\vec{x}_i)$ $(i = 1, \ldots, N)$ and $p(\vec{y}_j)$ $(j = 1, \ldots, M)$. It is obvious that

$$\sum_{i=1}^{N} \sum_{j=1}^{M} p(\vec{x}_i, \vec{y}_j) = 1 \tag{1}$$

$$\sum_{i=1}^{N} p(\vec{x}_i) = 1 \tag{2}$$

$$\sum_{j=1}^{M} p(\vec{y}_j) = 1. \tag{3}$$

Definition 1.

Entropy in terms of a group of discrete random variables is defined as

$$H(\vec{X}) = -\sum_{i=1}^{N} p(\vec{x}_i) \ln(p(\vec{x}_i)). \tag{4}$$

Consider that $0 \leq p(\vec{x}_i) \leq 1$ for $i = 1, 2, \ldots, N$, $H(\vec{X})$ in (4) is greater than or equal to zero; that is,

$$H(\vec{X}) \geq 0. \tag{5}$$

Thus, entropy can be thought as an extension of the determinant of the covariance matrix of $\vec{X}$. For the case of a single variable X, entropy is an extension of the variance of X.

Definition 2.

Conditional entropy of $\vec{X}$ given $\vec{Y}$ is defined by

$$H(\vec{X}|\vec{Y}) = -\sum_{i=1}^{N} \sum_{j=1}^{M} p(\vec{x}_i, \vec{y}_j) \ln(p(\vec{x}_i|\vec{y}_j)), \tag{6}$$

where $p(\vec{x}_i, \vec{y}_j)$ is the joint distribution of $(\vec{x}, \vec{y})$, and $p(\vec{x}_i|\vec{y}_j)$ is the conditional probability of $\vec{x}_i$ given $\vec{y}_j$.

Similar to (5),

$$H(\vec{X}|\vec{Y}) \geq 0. \tag{7}$$

Conditional entropy $H(\vec{X}|\vec{Y})$ measures the conditional uncertainty of $\vec{X}$ given $\vec{Y}$. In the case of $N = M = 1$, conditional entropy is analogous to the residual variance of a random variable X given another random variable Y. It is plausible that conditioning on $\vec{Y}$ should not increase the uncertainty about $\vec{X}$. This fact is expressed by the following theorem.

Theorem 1.

$$H(\vec{X}|\vec{Y}) \leq H(\vec{X})\ , with\ equality\ if\ and\ only\ if\ \vec{X}\ and\ \vec{Y}\ are\ independent. \tag{8}$$

In (8), equality is equivalent to the independence condition. This implies that entropy $H(\vec{X})$ and conditional entropy $H(\vec{X}|\vec{Y})$ contain information about the relationship or dependence between $\vec{X}$ and $\vec{Y}$. In past decades, (conditional) entropy had been explored as a means for measuring the association among random variables. If X and Y are both categorical variables (so that the joint frequency distribution for X and Y forms a two-way contingency table), Goodman and Kruskal (1954) proposed a number of methods in probabilistic terms to measure the degree of association between the two categorical random variables, which are represented by the raws and columns, respectively. Thiel (1970) proposed what he called the uncertainty coefficient (U), which is defined by the ratio of difference between the entropy of X and the conditional entropy of X given Y (information) to the entropy of X. This quantity is also described in Agresti (2002, pp. 56–57) and is available in both SAS and SPSS as part of the analysis of two-way contingency tables. Haberman (1980, 1982) also discussed the measures of associations and U in the context of analyzing multinomial response data. In this paper, a set of measures for the dependence of one group of random variables on another group of random variables is defined in terms of *entropy* and *conditional entropy*. The properties of these measures are explained and studied with the mathematical proofs.

2. K-Dependence Coefficient

First, define a new measure, which is called the *K-dependence coefficient* or *dependence coefficient*, for measuring the degree of dependence between two groups of random variables. Then, several properties of the K-dependence coefficient will be explored.

Definition 3.

$$K(\vec{X} : \vec{Y}) = \frac{H(\vec{X}) - H(\vec{X}|\vec{Y})}{H(\vec{X})}, \tag{9}$$

where

$H(\vec{X})$ is the entropy of $\vec{X}$, and

$H(\vec{X}|\vec{Y})$ is the conditional entropy of $\vec{X}$ given $\vec{Y}$.

$K(\vec{X} : \vec{Y})$ in (9) is called the K-dependence coefficient, or dependence coefficient, of $\vec{X}$ to $\vec{Y}$, or $\vec{X}$ dependence on $\vec{Y}$.

In Definition 3, the K-dependence coefficient is defined by the ratio of the difference between the entropy of $\vec{X}$ and the conditional entropy of $\vec{X}$ given $\vec{Y}$, and the entropy of $\vec{X}$. The difference between the entropy of $\vec{X}$ and the conditional entropy of $\vec{X}$ given $\vec{Y}$ actually measures the degree to which $\vec{X}$ depends on $\vec{Y}$, or $\vec{Y}$ depends on $\vec{X}$. With the K-dependence coefficient, the scale of this dependence measure is normalized using the entropy $H(\vec{X})$. The entropy $H(\vec{Y})$ can also be used for this normalization. The interpretations of these two normalizations are different and will be discussed later on.

Property 1.

The K-dependence coefficient ranges between 0 and 1 when $H(\vec{X}) > 0$; that is,

$$0 \le K(\vec{X} : \vec{Y}) \le 1. \tag{10}$$

Proof: This is obvious from (5) and Theorem 1.

Property 1 shows that $K(\vec{X} : \vec{Y})$ is a normalized measurement with the scale between 0 and

1. The following theorems look into $K(\vec{X}:\vec{Y})$ when its value is 0 or 1.

Theorem 2.

$$K(\vec{X}:\vec{Y}) = 0 \; if \; and \; only \; if \; \vec{X} \; and \; \vec{Y} \; are \; independent \; from \; each \; other. \tag{11}$$

Proof: From (9), it is obvious that

$$K(\vec{X}:\vec{Y}) = 0, \; if \; and \; only \; if \; H(\vec{X}|\vec{Y}) = H(\vec{X}).$$

According to Theorem 1,

$$H(\vec{X}|\vec{Y}) = H(\vec{X}) \; if \; and \; only \; if \; \vec{X} \; and \; \vec{Y} \; are \; independent.$$

Therefore,

$$K(\vec{X}:\vec{Y}) = 0 \; if \; and \; only \; if \; \vec{X} \; and \; \vec{Y} \; are \; independent \; from \; each \; other.$$

Because the independence of $\vec{X}$ from $\vec{Y}$ is equivalent to the independence of $\vec{Y}$ from $\vec{X}$, the following property is obvious.

Property 2.

$$K(\vec{X}:\vec{Y}) = 0 \;\; if \; and \; only \; if \;\; K(\vec{Y}:\vec{X}) = 0. \tag{12}$$

Proof: It is obvious from Theorem 2.

Theorem 2 looks into the case of $K(\vec{X}:\vec{Y}) = 0$, which is equivalent to independence between $\vec{X}$ and $\vec{Y}$. The following theorem looks into the case of $K(\vec{X}:\vec{Y}) = 1$.

Theorem 3.

Assume that $p(\vec{x}_i) > 0$ and $p(\vec{y}_j) > 0 \quad (i = 1, 2, \ldots, N; j = 1, 2, \ldots, M)$. Then

$$K(\vec{X}:\vec{Y}) = 1$$

if and only if for each $\vec{y}_j$, $\exists \quad \vec{x}_{i_j}$ such that $p(\vec{X} = \vec{x}_{i_j} | \vec{Y} = \vec{y}_j) = 1$ and $p(\vec{X} = \vec{x}_k | \vec{Y} = \vec{y}_j) = 0$, where $k \neq i_j$.

Proof: First, necessity is obvious because $H(\vec{X}|\vec{Y}) = 0$, from Definition 2, also noting that $\mathit{0} \ln(\mathit{0}) = \mathit{0}$. Therefore, $K(\vec{X} : \vec{Y}) = 1$ by Definition 3.

Now, assume that $K(\vec{X} : \vec{Y}) = 1$. From Definition 3,

$$H(\vec{X}|\vec{Y}) = 0; \tag{13}$$

that is,

$$0 = H(\vec{X}|\vec{Y}) = -\sum_{j=1}^{M} p(\vec{y}_j) \sum_{i=1}^{N} p(\vec{x}_i|\vec{y}_j) \ln(p(\vec{x}_i|\vec{y}_j) = -\sum_{j=1}^{M} \sum_{i=1}^{N} p(\vec{x}_i, \vec{y}_j) \ln(p(\vec{x}_i|\vec{y}_j)).$$

Therefore,

$$\sum_{j=1}^{M} \sum_{i=1}^{N} p(\vec{x}_i, \vec{y}_j) \ln(p(\vec{x}_i|\vec{y}_j)) = 0. \tag{14}$$

Also, it is noted that

$$\begin{aligned} &0 \leq p(\vec{x}_i, \vec{y}_j) \leq 1, and \\ &0 \leq p(\vec{x}_i|\vec{y}_j) \leq 1. \end{aligned}$$

Therefore,

$$p(\vec{x}_i, \vec{y}_j) \ln(p(\vec{x}_i|\vec{y}_j)) \leq 0 \;\; for\ all\ i\ and\ j. \tag{15}$$

Combining (14) and (15),

$$p(\vec{x}_i, \vec{y}_j) \ln(p(\vec{x}_i|\vec{y}_j)) = 0 \;\; for\ all\ i\ and\ j. \tag{16}$$

For each given $\vec{y}_j$,

$$\sum_{i=1}^{N} p(\vec{x}_i, \vec{y}_j) = p(\vec{y}_j) > 0, and \tag{17}$$

$$p(\vec{x}_i, \vec{y}_j) \geq 0 \;\; for \; all \; i. \tag{18}$$

Here, $p(\vec{y}_j) > 0$ is the assumption. From (17), there exists at least one i_j, such that

$$p(\vec{x}_{i_j}, \vec{y}_j) > 0. \tag{19}$$

From (16),

$$p(\vec{x}_{i_j}, \vec{y}_j) \ln(p(\vec{x}_{i_j} | \vec{y}_j)) = 0. \tag{20}$$

Combining (19) and (20),

$$\ln(p(\vec{x}_{i_j} | \vec{y}_j)) = 0. \tag{21}$$

Therefore,

$$p(\vec{x}_{i_j} | \vec{y}_j) = 1. \tag{22}$$

Also,

$$\sum_{i=1}^{N} p(\vec{x}_i | \vec{y}_j) = 1. \tag{23}$$

Combining (22) and (23),

$$p(\vec{x}_i | \vec{y}_j) = 0 \;\; for \; all \; i \; \neq \; i_j. \tag{24}$$

Therefore, (22) and (24) are the proof for sufficiency.

Theorems 2 and 3 show that $K(\vec{X} : \vec{Y})$ measures a total dependence or total relationship, not only the linear part of this dependence or relationship, between $\vec{X}$ and $\vec{Y}$. More specifically, $K(\vec{X} : \vec{Y})$ actually measures the $\vec{X}$'s dependence on $\vec{Y}$. Generally speaking, this dependence is not necessarily symmetric; that is, $K(\vec{X} : \vec{Y})$ may or may not be the same as $K(\vec{Y} : \vec{X})$. This is a different situation from linear correlation, where the relationship is always symmetric. Because

$K(\vec{X}:\vec{Y})$ measures $\vec{X}$'s dependence on $\vec{Y}$, while $K(\vec{Y}:\vec{X})$ measures $\vec{Y}$'s dependence on $\vec{X}$, $K(\vec{X}:\vec{Y})$ has no property of symmetry.

Next is a proof for a K-dependence coefficient inequality.

Property 3.

$$K(\vec{X}:\vec{Y}) \leq K(\vec{X}:\vec{Y},\vec{Z}), \tag{25}$$

with equality if and only if $\vec{X}$ *and* $\vec{Z}$ *are conditionally independent given* $\vec{Y}$.

Proof: Assume $\vec{X}, \vec{Y}$, and $\vec{Z}$ have N, M, and L states, respectively, which are denoted by $\{\vec{x}_1, \ldots, \vec{x}_N\}$, $\{\vec{y}_1, \ldots, \vec{y}_M\}$, and $\{\vec{z}_1, \ldots, \vec{z}_L\}$.

First,

$$\begin{aligned}
H(\vec{X},\vec{Z}|\vec{Y}) &= -\sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{k=1}^{L} p(\vec{x}_i,\vec{y}_j,\vec{z}_k)\ln(p(\vec{x}_i,\vec{z}_k|\vec{y}_j)) \\
&= -\sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{k=1}^{L} p(\vec{x}_i,\vec{y}_j,\vec{z}_k)\ln(p(\vec{z}_k|\vec{y}_j)p(\vec{x}_k|\vec{y}_j,\vec{z}_k)) \\
&= -\sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{k=1}^{L} p(\vec{x}_i,\vec{y}_j,\vec{z}_k)\ln(p(\vec{z}_k|\vec{y}_j)) - \sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{k=1}^{L} p(\vec{x}_i,\vec{y}_j,\vec{z}_k)\ln(p(\vec{x}_k|\vec{y}_j,\vec{z}_k)) \\
&= H(\vec{Z}|\vec{Y}) + H(\vec{X}|\vec{Y},\vec{Z}).
\end{aligned}$$

Therefore,

$$H(\vec{X},\vec{Z}|\vec{Y}) = H(\vec{Z}|\vec{Y}) + H(\vec{X}|\vec{Y},\vec{Z}). \tag{26}$$

Second, it has been known that

$$\ln(x) \leq x-1, \;\; \textit{with equality if and only if} \;\; x=1.$$

Therefore, for each i, j, and k,

$$\ln(\frac{p(\vec{x}_i|\vec{y}_j)p(\vec{z}_k|\vec{y}_j)}{p(\vec{x}_i,\vec{z}_k|\vec{y}_j)}) \leq \frac{p(\vec{x}_i|\vec{y}_j)p(\vec{z}_k|\vec{y}_j)}{p(\vec{x}_i,\vec{z}_k|\vec{y}_j)} - 1, \tag{27}$$

$$with\ \ equality\ \ if\, and\, only\, if\ \ p(\vec{x}_i,\vec{z}_k|\vec{y}_j) = p(\vec{x}_i|\vec{y}_j)p(\vec{z}_k|\vec{y}_j).$$

Multiply (27) by $p(\vec{x}_i,\vec{z}_k|\vec{y}_j)$ and sum over i and k to obtain

$$\sum_{i=1}^{N}\sum_{k=1}^{L} p(\vec{x}_i,\vec{z}_k|\vec{y}_j)\ln(\frac{p(\vec{x}_i|\vec{y}_j)p(\vec{z}_k|\vec{y}_j)}{p(\vec{x}_i,\vec{z}_k|\vec{y}_j)}) \leq \sum_{i=1}^{N}\sum_{k=1}^{L} p(\vec{x}_i,\vec{z}_k|\vec{y}_j)(\frac{p(\vec{x}_i|\vec{y}_j)p(\vec{z}_k|\vec{y}_j)}{p(\vec{x}_i,\vec{z}_k|\vec{y}_j)} - 1)$$

$$= \sum_{i=1}^{N}\sum_{k=1}^{L}(p(\vec{x}_i|\vec{y}_j)p(\vec{z}_k|\vec{y}_j) - p(\vec{x}_i,\vec{z}_k|\vec{y}_j)) = 1 - 1 = 0, \tag{28}$$

$$with\ \ equality\ \ if\, and\, only\, if\ \ p(\vec{x}_i,\vec{z}_k|\vec{y}_j) = p(\vec{x}_i|\vec{y}_j)p(\vec{z}_k|\vec{y}_j)\ \ for\ \ all\ \ i,j,\ \ and\ \ k.$$

From (28),

$$\sum_{i=1}^{N}\sum_{k=1}^{L} p(\vec{x}_i,\vec{z}_k|\vec{y}_j)\ln(p(\vec{x}_i|\vec{y}_j)) + \sum_{i=1}^{N}\sum_{k=1}^{L} p(\vec{x}_i,\vec{z}_k|\vec{y}_j)\ln(p(\vec{z}_k|\vec{y}_j))$$
$$\leq \sum_{i=1}^{N}\sum_{k=1}^{L} p(\vec{x}_i,\vec{z}_k|\vec{y}_j)\ln(p(\vec{x}_i,\vec{z}_k|\vec{y}_j)), \tag{29}$$

$$with\ \ equality\ \ if\, and\, only\, if\ \ p(\vec{x}_i,\vec{z}_k|\vec{y}_j) = p(\vec{x}_i|\vec{y}_j)p(\vec{z}_k|\vec{y}_j)\ \ for\ \ all\ \ j.$$

Multiply (29) by $-p(\vec{y}_j)$ and sum over j to obtain

$$-\sum_{j=1}^{M} p(\vec{y}_j)\sum_{i=1}^{N}\sum_{k=1}^{L} p(\vec{x}_i,\vec{z}_k|\vec{y}_j)\ln(p(\vec{x}_i|\vec{y}_j)) - \sum_{j=1}^{M} p(\vec{y}_j)\sum_{i=1}^{N}\sum_{k=1}^{L} p(\vec{x}_i,\vec{z}_k|\vec{y}_j)\ln(p(\vec{z}_k|\vec{y}_j))$$
$$\geq -\sum_{j=1}^{M} p(\vec{y}_j)\sum_{i=1}^{N}\sum_{k=1}^{L} p(\vec{x}_i,\vec{z}_k|\vec{y}_j)\ln(p(\vec{x}_i,\vec{z}_k|\vec{y}_j)), \tag{30}$$

$$with\ \ equality\ \ if\, and\, only\, if\ \ p(\vec{x}_i,\vec{z}_k|\vec{y}_j) = p(\vec{x}_i|\vec{y}_j)p(\vec{z}_k|\vec{y}_j)\ \ for\ \ all\ \ i,j,\ \ and\ \ k;$$

that is,

$$H(\vec{X}, \vec{Z}|\vec{Y}) \leq H(\vec{X}|\vec{Y}) + H(\vec{Z}|\vec{Y}), \tag{31}$$

with equality if and only if $p(\vec{x}_i, \vec{z}_k|\vec{y}_j) = p(\vec{x}_i|\vec{y}_j)p(\vec{z}_k|\vec{y}_j)$ *for all* $i, j,$ *and* k.

From (26) and (31),

$$H(\vec{X}|\vec{Y}, \vec{Z}) = H(\vec{X}, \vec{Z}|\vec{Y}) - H(\vec{Z}|\vec{Y})$$
$$\leq H(\vec{X}|\vec{Y}) + H(\vec{Z}|\vec{Y}) - H(\vec{Z}|\vec{Y}) = H(\vec{X}|\vec{Y}).$$

Therefore,

$$H(\vec{X}|\vec{Y}, \vec{Z}) \leq H(\vec{X}|\vec{Y}), \tag{32}$$

with equality if and only if $\vec{X}$ *and* $\vec{Z}$ *are conditionally independent given* $\vec{Y}$.

From Definition 3,

$$K(\vec{X} : \vec{Y}) \leq K(\vec{X} : \vec{Y}, \vec{Z}),$$

with equality if and only if $\vec{X}$ *and* $\vec{Z}$ *are conditionally independent given* $\vec{Y}$.

Property 3 is also true for the case of multivariables; that is:

Property 4.

$$K(\vec{X} : \vec{Y}_1, \ldots, \vec{Y}_p) \leq K(\vec{X} : \vec{Y}_1, \ldots, \vec{Y}_p, \vec{Y}_{p+1}),$$

with equality if and only if $\vec{X}$ *and* $\vec{Y}_{p+1}$ *are conditionally independent given* $\vec{Y}_1, \ldots, \vec{Y}_p$.

3. Partial and Semipartial K-Dependence Coefficients

Section 2 defines the K-dependence coefficient, which measures the degree to which a group of random variables depends on another group of random variables. This section looks into a conditional K-dependence coefficient, which measures the degree to which one group of random variables depends on a second group of random variables, given a third group of random variables. This conditional K-dependence coefficient is called the *partial K-dependence coefficient*, or the *partial dependence coefficient*.

Let $\vec{Z} = (z_1, z_2, \ldots, z_l)^T$ be a group of discrete random variables with L possible states, which are denoted by $\{\vec{z}_1, \ldots, \vec{z}_L\}$. The probability of each state is denoted by $p(\vec{z}_k)$ $(k = 1, 2, \ldots, L)$.

Definition 4.

$$K(\vec{X} : \vec{Y}|\vec{Z}) = \frac{H(\vec{X}|\vec{Z}) - H(\vec{X}|\vec{Y}, \vec{Z})}{H(\vec{X}|\vec{Z})}, \tag{33}$$

where

$H(\vec{X}|\vec{Z})$ is the conditional entropy of $\vec{X}$ given $\vec{Z}$, and

$H(\vec{X}|\vec{Y}, \vec{Z})$ is the conditional entropy of $\vec{X}$ given $\vec{Y}$ and $\vec{Z}$.

$K(\vec{X} : \vec{Y}|\vec{Z})$ defined in (33) is called the partial *K-dependence coefficient*, or the *partial dependence coefficient* of $\vec{X}$ to $\vec{Y}$, or $\vec{X}$ dependence on $\vec{Y}$ given $\vec{Z}$.

Theorem 4.

$$K(\vec{X} : \vec{Y}|\vec{Z}) = 0 \; if \; and \; only \; if \; \vec{X} \; and \; \vec{Y} \; are \; conditionally \; independent \; given \; \vec{Z}. \tag{34}$$

Proof: From (32), it is obvious that

$$H(\vec{X}|\vec{Z}) = H(\vec{X}|\vec{Y}, \vec{Z}) \tag{35}$$

if and only if $\vec{X}$ and $\vec{Y}$ are conditionally independent given $\vec{Z}$.

Also, from Definition 4, (35) is true if and only if

$$K(\vec{X} : \vec{Y}|\vec{Z}) = 0. \tag{36}$$

Therefore, (36) is true if and only if $\vec{X}$ and $\vec{Y}$ are conditionally independent given $\vec{Z}$.

Theorem 5.

Assume that $p(\vec{x}_i) > 0$, $p(\vec{y}_j) > 0$ and $p(\vec{z}_k) > 0 \quad (i = 1, 2, \ldots, N;\ j = 1, 2, \ldots, M;\ k = 1, 2, \ldots, L)$ Then

$$K(\vec{X} : \vec{Y}|\vec{Z}) = 1$$

if and only if for each $\vec{y}_j$ and $\vec{z}_k$, $\exists \quad \vec{x}_{i_{j,k}}$, such that $p(\vec{X} = \vec{x}_{i_{j,k}}|\vec{Y} = \vec{y}_j, \vec{Z} = \vec{z}_k) = 1$ and $p(\vec{X} = \vec{x}_l|\vec{Y} = \vec{y}_j, \vec{Z} = \vec{z}_k) = 0$, where $l \neq i_{j,k}$.

Proof: From Definition 4,

$$K(\vec{X} : \vec{Y}|\vec{Z}) = 1 \tag{37}$$

if and only if

$$H(\vec{X}|\vec{Y}, \vec{Z}) = 0. \tag{38}$$

Now, define

$$\vec{\Phi} = (\vec{Y}, \vec{Z})^T. \tag{39}$$

$\vec{\Phi}$ is also a vector of random variables with $M \times L$ states:

$$\vec{\phi}_{j,k} = (\vec{y}_j, \vec{z}_k)^T. \tag{40}$$

Then (38) becomes

$$H(\vec{X}|\vec{\Phi}) = 0. \tag{41}$$

By applying Theorem 3 to $\vec{\Phi}$ instead of $\vec{Y}$, it can be shown that (41) is true if and only if for each $\vec{\phi}_{j,k}$, $\exists \quad \vec{x}_{i_{j,k}}$ such that $p(\vec{X} = \vec{x}_{i_{j,k}}|\vec{\Phi} = \vec{\phi}_{j,k}) = 1$ and $p(\vec{X} = \vec{x}_l|\vec{\Phi} = \vec{\phi}_{j,k}) = 0$, where $l \neq i_{j,k}$.

This is the same as the statement that for each $\vec{y}_j$ and $\vec{z}_k$, $\exists \quad \vec{x}_{i_{j,k}}$ such that $p(\vec{X} = \vec{x}_{i_{j,k}}|\vec{Y} = \vec{y}_j, \vec{Z} = \vec{z}_k) = 1$ and $p(\vec{X} = \vec{x}_l|\vec{Y} = \vec{y}_j, \vec{Z} = \vec{z}_k) = 0$, where $l \neq i_{j,k}$.

Also, from Definition 4, (41) is true if and only if

$$K(\vec{X}:\vec{Y}|\vec{Z}) = 1.$$

This proves Theorem 5.

Similar to Property 3, the partial K-dependence coefficient satisfies the following inequality.

Property 5.

$$K(\vec{X}:\vec{Y}_1|\vec{Z}) \leq K(\vec{X}:\vec{Y}_1,\vec{Y}_2|\vec{Z}), \tag{42}$$

with equality if and only if $\vec{X}$ *and* $\vec{Y}_2$ *are conditionally independent given* $\vec{Y}_1$ *and* $\vec{Z}$.

Proof: In the same way as (26) and (31), it can be proved that

$$H(\vec{X},\vec{Y}_2|\vec{Y}_1,\vec{Z}) = H(\vec{Y}_2|\vec{Y}_1,\vec{Z}) + H(\vec{X}|\vec{Y}_1,\vec{Y}_2,\vec{Z}),\ and \tag{43}$$

$$H(\vec{X},\vec{Y}_2|\vec{Y}_1,\vec{Z}) \leq H(\vec{X}|\vec{Y}_1,\vec{Z}) + H(\vec{Y}_2|\vec{Y}_1,\vec{Z}), \tag{44}$$

with equality if and only if $\vec{X}$ *and* $\vec{Y}_2$ *are conditionally independent, given* $\vec{Y}_1$ *and* $\vec{Z}$.

From (43) and (44),

$$\begin{aligned} H(\vec{X}|\vec{Y}_1,\vec{Y}_2,\vec{Z}) &= H(\vec{X},\vec{Y}_2|\vec{Y}_1,\vec{Z}) - H(\vec{Y}_2|\vec{Y}_1,\vec{Z}) \\ &\leq H(\vec{X}|\vec{Y}_1,\vec{Z}) + H(\vec{Y}_2|\vec{Y}_1,\vec{Z}) - H(\vec{Y}_2|\vec{Y}_1,\vec{Z}) = H(\vec{X}|\vec{Y}_1,\vec{Z}). \end{aligned}$$

Therefore,

$$H(\vec{X}|\vec{Y}_1,\vec{Y}_2,\vec{Z}) \leq H(\vec{X}|\vec{Y}_1,\vec{Z}),$$

with equality if and only if $\vec{X}$ *and* $\vec{Y}_2$ *are conditionally independent, given* $\vec{Y}_1$ *and* $\vec{Z}$.

From Definition 4,

$$K(\vec{X} : \vec{Y}_1|\vec{Z}) \leq K(\vec{X} : \vec{Y}_1, \vec{Y}_2|\vec{Z}),$$

with equality if and only if $\vec{X}$ *and* $\vec{Y}_1$ *are conditionally independent, given* $\vec{Y}_1$ *and* $\vec{Z}$.

Definition 5.

$$K(\vec{X} : (\vec{Y}|\vec{Z})) = \frac{H(\vec{X}|\vec{Z}) - H(\vec{X}|\vec{Y}, \vec{Z})}{H(\vec{X})}, \tag{45}$$

where

$H(\vec{X}|\vec{Z})$ is the conditional entropy of $\vec{X}$ given $\vec{Z}$, and

$H(\vec{X}|\vec{Y}, \vec{Z})$ is the conditional entropy of $\vec{X}$ given $\vec{Y}$ and $\vec{Z}$.

$K(\vec{X} : (\vec{Y}|\vec{Z}))$ in (45) is called the *semipartial K-dependence coefficient*, or the *semipartial dependence coefficient* of $\vec{X}$ to $\vec{Y}$, or $\vec{X}$ dependence on conditional $\vec{Y}$ given $\vec{Z}$.

The semipartial K-dependence coefficient differs from the partial K-dependence coefficient by a factor of $H(\vec{X}|\vec{Z})/H(\vec{X})$; that is,

$$K(\vec{X} : (\vec{Y}|\vec{Z})) = K(\vec{X} : \vec{Y}|\vec{Z}) \times \frac{H(\vec{X}|\vec{Z})}{H(\vec{X})}, \; and \tag{46}$$

Therefore,

$$K(\vec{X} : (\vec{Y}|\vec{Z})) \leq K(\vec{X} : \vec{Y}|\vec{Z}).$$

Theorem 6.

$$1 - K(\vec{X} : \vec{Y}_1, \ldots, \vec{Y}_n) = (1 - K(\vec{X} : \vec{Y}_1)) \times (1 - K(\vec{X} : \vec{Y}_2|\vec{Y}_1)) \tag{47}$$

$$\times \cdots \times (1 - K(\vec{X} : \vec{Y}_n|\vec{Y}_1, \ldots, \vec{Y}_{n-1})).$$

Proof:

$$1 - K(\vec{X} : \vec{Y}_1, \ldots, \vec{Y}_n) = \frac{H(\vec{X}|\vec{Y}_1, \ldots, \vec{Y}_n)}{H(\vec{X})}$$

$$= \frac{H(\vec{X}|\vec{Y}_1)}{H(\vec{X})} \times \frac{H(\vec{X}|\vec{Y}_1, \vec{Y}_2)}{H(\vec{X}|\vec{Y}_1)} \times \frac{H(\vec{X}|\vec{Y}_1, \vec{Y}_2, \vec{Y}_3)}{H(\vec{X}|\vec{Y}_1, \vec{Y}_2)} \times \cdots \times \frac{H(\vec{X}|\vec{Y}_1, \ldots, \vec{Y}_{n-1}, \vec{Y}_n)}{H(\vec{X}|\vec{Y}_1, \ldots, \vec{Y}_{n-1})}$$

$$= (1 - K(\vec{X} : \vec{Y}_1)) \times (1 - K(\vec{X} : \vec{Y}_2|\vec{Y}_1)) \times \cdots \times (1 - K(\vec{X} : \vec{Y}_n|\vec{Y}_1, \ldots, \vec{Y}_{n-1})).$$

Theorem 6 provides a decomposition of the K-dependence coefficient in terms of the partial K-dependence coefficient. The following theorem expresses the partial K-dependence coefficient in terms of the K-dependence coefficient.

Theorem 7.

$$1 - K(\vec{X} : \vec{Y}|\vec{Z}) = \frac{1 - K(\vec{X} : \vec{Y}, \vec{Z})}{1 - K(\vec{X} : \vec{Z})}. \tag{48}$$

Proof:

$$\frac{H(\vec{X}|\vec{Y}, \vec{Z})}{H(\vec{X}|\vec{Z})} = \frac{H(\vec{X}|\vec{Y}, \vec{Z})}{H(\vec{X}|\vec{Z})}$$

$$\Longleftrightarrow$$

$$1 - \frac{H(\vec{X}|\vec{Z}) - H(\vec{X}|\vec{Y}, \vec{Z})}{H(\vec{X}|\vec{Z})} = \frac{1 - \frac{H(\vec{X}) - H(\vec{X}|\vec{Y}, \vec{Z})}{H(\vec{X})}}{1 - \frac{H(\vec{X}) - H(\vec{Y}|\vec{Z})}{H(\vec{X})}}$$

$$\Longleftrightarrow$$

$$1 - K(\vec{X} : \vec{Y}|\vec{Z}) = \frac{1 - K(\vec{X} : \vec{Y}, \vec{Z})}{1 - K(\vec{X} : \vec{Z})}.$$

Theorem 8.

$$K(\vec{X} : \vec{Y}|\vec{Z}) = 0 \;\; if \; and \; only \; if \;\; K(\vec{X} : (\vec{Y}|\vec{Z})) = 0 \tag{49}$$

Proof: This is obvious from Definition 4 and Definition 5.

Theorem 8 shows that the independence between $\vec{X}$ and $\vec{Y}$, with the condition of $\vec{Z}$ for both $\vec{X}$ and $\vec{Y}$, is equivalent to the independence between $\vec{X}$ and $\vec{Y}$, with the condition of $\vec{Z}$ only for $\vec{Y}$.

Theorem 9.

$$K(\vec{X} : \vec{Y}_1, \vec{Y}_2, \ldots, \vec{Y}_n) = \tag{50}$$

$$K(\vec{X} : \vec{Y}_1) + K(\vec{X} : (\vec{Y}_2|\vec{Y}_1)) + K(\vec{X} : (\vec{Y}_3|\vec{Y}_1, \vec{Y}_2)) + \cdots + K(\vec{X} : (\vec{Y}_n|\vec{Y}_1, \ldots, \vec{Y}_{n-1})).$$

Proof:

$$K(\vec{X} : \vec{Y}_1, \vec{Y}_2, \ldots, \vec{Y}_n) = \frac{H(\vec{X}) - H(\vec{X}|\vec{Y}_1, \vec{Y}_2, \ldots, \vec{Y}_n)}{H(\vec{X})}$$

$$= \frac{1}{H(\vec{X})}[(H(\vec{X}) - H(\vec{X}|\vec{Y}_1)) + (H(\vec{X}|\vec{Y}_1) - H(\vec{X}|\vec{Y}_1, \vec{Y}_2)) + \cdots +$$

$$(H(\vec{X}|\vec{Y}_1, \ldots, \vec{Y}_{n-1}) - H(\vec{X}|\vec{Y}_1, \ldots, \vec{Y}_{n-1}, \vec{Y}_n))]$$

$$= K(\vec{X} : \vec{Y}_1) + K(\vec{X} : (\vec{Y}_2|\vec{Y}_1)) + K(\vec{X} : (\vec{Y}_3|\vec{Y}_1, \vec{Y}_2)) + \cdots + K(\vec{X} : (\vec{Y}_n|\vec{Y}_1, \ldots, \vec{Y}_{n-1})).$$

In (50), the total degree of $\vec{X}$'s dependence on $\vec{Y}_1, \ldots, \vec{Y}_n$ is actually the summation of the degrees of the $\vec{X}$'s dependencies on each $\vec{Y}_i$'s unique contribution to $\vec{X}$, with common contribution removed.

Theorem 10.

$$K(\vec{X} : \vec{Y}) = K(\vec{Y} : \vec{X}) \ \ if\ and\ only\ if \ \ H(\vec{X}) = H(\vec{Y}),\ and \tag{51}$$

$$K(\vec{X} : \vec{Y}|\vec{Z}) = K(\vec{Y} : \vec{X}|\vec{Z}) \ \ if\ and\ only\ if \ \ H(\vec{X}|\vec{Z}) = H(\vec{Y}|\vec{Z}). \tag{52}$$

Proof: (51) is obvious from Definition 3 and the following the well-known equation for mutual information:

$$H(\vec{X}) - H(\vec{X}|\vec{Y}) = H(\vec{Y}) - H(\vec{Y}|\vec{X}). \tag{53}$$

$H(\vec{X}) - H(\vec{X}|\vec{Y})$ in (53) is called mutual information among $\vec{X}$ and $\vec{Y}$ and denoted by $I(\vec{X}, \vec{Y})$.

Equation (53) is also true if one substitutes $\vec{X}$ and $\vec{Y}$ with $\vec{X}|\vec{Z}$ and $\vec{Y}|\vec{Z}$, respectively. Here, $\vec{X}|\vec{Z}$ and $\vec{Y}|\vec{Z}$ are the random vectors defined from random vectors $\vec{X}$ and $\vec{Y}$, given condition $\vec{Z}$. Thus,

$$H(\vec{X}|\vec{Z}) - H(\vec{X}|\vec{Y}, \vec{Z}) = H(\vec{Y}|\vec{Z}) - H(\vec{Y}|\vec{X}, \vec{Z}). \tag{54}$$

Equation (54) plus Definition 4 are the proof of (52).

Theorem 10 shows that, in general, $K(\vec{X} : \vec{Y})$ may or may not be the same as $K(\vec{Y} : \vec{X})$. This is called asymmetry. This property of asymmetry does not hold with linear correlations because linear relationships are always symmetric. The K-dependence coefficient measures the degree to which a group of random variables depend on another group of random variables. This dependence may or may not be linear. Generally speaking, for nonlinear dependence, one should not expect the measure to be always symmetric.

4. Discussion

This paper has defined the K-dependence coefficient and provided the proofs of some properties. The idea behind the K-dependence coefficient in (9) is that it represents the conditional entropy normalized by the unconditional entropy. Actually, other versions of this measure have appeared in the statistical literature, which have been discussed in the introduction. In 1988, Burg and Lewis compared two kinds of measures of multivariate association, based on Wilks'Λ and the Bartlett-Nanda-Pillai trace criterion V in terms of properties of the univariate R^2. A unified set of derivations of the properties of R^2 is provided. These properties are also thought to be true for the K-dependence coefficient. Following the definition of the K-dependence coefficient, the partial K-dependence coefficient and the semipartial K-dependence coefficient are also defined (Charles Lewis, personnel communication, May 9, 2006). A set of the properties for (partial or semipartial) K-dependence coefficients is derived accordingly in Section 3.

For the K-dependence coefficient, there are at least two aspects that are different from linear correlation. First, the K-dependence coefficient measures the degree to which a group of random variables depend on another group of random variables. This dependency may or may not be linear. In fact, the dependence measured by the K-dependence coefficient includes both linear and nonlinear dependence among random variables. Secondly, the K-dependence coefficient may or may not be symmetric; that is, $K(\vec{X}, \vec{Y})$ may or may not be the same as $K(\vec{Y}, \vec{X})$. Linear dependence is always symmetric; but for nonlinear dependence, symmetry does not hold if and only if $H(\vec{X}) \neq H(\vec{Y})$, according to Theorem 10. Generally speaking, symmetry is always true for linear dependence, but asymmetry in general holds for nonlinear dependence.

Although the partial K-dependence coefficient and the semipartial K-dependence coefficients are defined separately, essentially, the semipartial K-dependence coefficient is a more general concept; that is, the partial K-dependence coefficient is a special case of the semipartial K-dependence coefficient. This is because

$$K(\vec{X} : \vec{Y} | \vec{Z}) = K((\vec{X}|\vec{Z}) : (\vec{Y}|\vec{Z})). \tag{55}$$

The right-hand side of (55) is actually defined from the definition of the semipartial K-dependence coefficient. If $\vec{Z}$ is independent from $\vec{X}$ and $\vec{Y}$, the left-hand side of (55) becomes the K-dependence coefficient. Therefore, the K-dependence coefficient and the partial K-dependence coefficient can be thought to be special cases of the semipartial K-dependence coefficient; that is, the K-dependence coefficient, the partial K-dependence coefficient, and the semipartial K-dependence coefficient actually are unified under the concept of the semipartial K-dependence coefficient. Therefore, any conclusions for the semipartial K-dependence coefficient are also relevant for the K-dependence coefficient and the partial K-dependence coefficient.

References

Agresti, A. (2002). *Categorical data analysis.* New York: John Wiley.

Burg, W., & Lewis, C. (1988, March). Some properties of two measures of multivariate association. *Psychometrika, 53*, 109–122.

Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross classification. *Journal of the American Statistical Association*, *49*, 732–64.

Haberman, S. J. (1980). Measures of association. In S. Kotz & N. L. Johnson (Eds.), *Encyclopedia of statistical sciences* (vol. 1, pp. 130–137). New York: John Wiley.

Haberman, S. J. (1982). Analysis of dispersion of multinomial responses. *Journal of the American Statistical Association*, *77*, 568–580.

Theil, H. (1970). On the estimation of relationships involving qualitative variables. *American Journal of Sociology*, *76*, 103–154.